Bayesian Model Selection for Complex Geological Structures Using Polynomial Chaos Proxy

Bayesian Model Selection for Complex Geological Structures Using Polynomial Chaos Proxy

Hamid Bazargan[a,*], Mike Christie[a]

[a]*Heriot Watt University, Edinburgh, UK, EH14 4AS*

Abstract

Different interpretation of sedimentary environments lead to scenario uncertainty where the prior reservoir model has a high level of discrete uncertainty. In a real field application, the scenario uncertainty has a considerable effect on flow response uncertainty and makes the uncertainty quantification problem highly nonlinear. We use clustering methods to address the scenario uncertainty. Our approach to cluster analysis is based on the posterior probabilities of models, known as "Bayesian Model Selection". Accordingly, we integrate over all possible parameters in each scenario with respect to their corresponding priors to give the measure of how well a model is supported by observations.

We propose a cluster-based reduced terms polynomial chaos proxy to efficiently estimate the posterior probability density function under each cluster and calculate the posterior probability of each model. We demonstrate that the convergence rate of the reduced terms polynomial chaos proxy is significantly improved under each cluster comparing to the non-clustered case. We apply the proposed cluster-based polynomial chaos proxy framework to study the plausibility of three training images based on different geological interpretation of the second layer of synthetic Stanford VI reservoir. We demonstrate that the proposed workflow can be efficiently used to calculate the posterior probability of each scenario and also sample from the posterior facies models within each scenario.

Keywords: Bayesian Parameter Estimation, Multiple Training Image, Bayesian Model Selection, Polynomial Chaos Expansion, Mixture Modelling

1. Introduction

Classification is a common practice in various fields of science. As Kendall said, "*one of the basic problems of science in reducing the world to order (or, if you prefer it, in imposing a manmade order on the complexity of things) is to classify*" [2, 3]. In general, classification can be defined as the clustering of objects based on their similar characteristics. For cluster analysis, various methods have been practiced in literature [3, 6, 7]. These methods generally

*Corresponding author, University of Warwick, Coventry, CV4 7AL, UK
Email address: hamidbazargan@gmail.com (Hamid Bazargan)

vary from mainly heuristic methods to methods based on statistical techniques. For example, one popular algorithm is hierarchical clustering, where based on an optimality criterion at each step two clusters are either merged (agglomerative) or divided (divisive) [8, 3]. Another well-known technique is the K-mean algorithm that relocates the observation between a predetermined number of clusters [9, 3]. The K-mean algorithm clusters data into K different groups by minimizing the total mean square error between the training samples and their representative cluster centroid [9]. Both of these methods require either a prior knowledge on the number of the clusters or a pre-determined optimality criterion to optimize the number of the clusters [3]. Essentially, all the cluster analysis techniques have been developed to determine the number and structures of the clusters [3]. These entail the determination of

1. similarity and separation of clusters,
2. shape (distribution) of underlying clusters,
3. relative sizes and compactness of clusters [3].

Mixture modeling is a probabilistic approach to clustering where datasets are assumed to be samples from a mixture of clusters with different probability density functions [10]. Then each individual sample does not exclusively belongs to a specific cluster. Hence, in the mixture modeling approach the issue of cluster analysis reduces to a single concern, that of model selection [10]. Considering that each combination of clusters results in a different overall distribution for the data, cluster analysis essentially concerns the problem of comparison among the number of possible models and the probability density function of the underlying clusters [10].

There are usually trade-offs between the number of clusters and the complexity of the underlying models. If a simpler model is used, then more clusters may be needed to provide an accurate representation of the data. If a more complex model is used, then data can be represented with a few clusters [3]. For example, to fit data from a single Gaussian cluster whose covariance is a single elongated ellipsoid, one requires more than one hyper-spherical Gaussian cluster.

In this work, we assume the geological understudy is highly nonlinear and complex, but can be effectively clustered into different scenarios. For example, we discuss the case that the permeability realizations can be generated using different training images. The objectives of clustering in our work are:

1. to make the problem less nonlinear under each cluster. Hence, the polynomial chaos proxy converges faster for each cluster and the overall posterior distribution can be estimated more efficiently using different proxies.
2. to compare the possibility of different scenarios given data.

Our approach to cluster analysis in mixture modeling is based on the posterior probabilities of models, known as the "Bayesian Model Selection". In statistics, the Bayesian model selection

was first introduced in the work of Jeffreys in 1939 [11] and pioneered by Cox [12], Akaike [13], Schwarz [14]. A rich review on the Bayesian model selection is presented by Clyde and George [15]. To illustrate the Bayesian model selection consider models $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_K$ with prior probabilities $P(\mathcal{M}_j),\ j = 1, \ldots, K$ (often assumed to be equal), then by Bayes's rule, the posterior probability of model $\mathcal{M}_j$ given data $\mathcal{D}$ is proportional to the probability of the data given model $\mathcal{M}_j$, times the model's prior probability, namely

$$P(\mathcal{M}_j|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M}_j)P(\mathcal{M}_j). \tag{1}$$

For the posterior calculation, the Markov Chain Monte Carlo (MCMC) has been successfully practiced in the Bayesian framework [16, 17, 36, 31, 20]. Raftery et al. examined application of the Bayesian model selection in five areas; genetics, sports, ecology, sociology, psychology [21, 22, 23]. Subsequently, in the reservoir modeling context, Gallagher et al. applied MCMC for the Bayesian inference to select the optimal reservoir model [24]. In large problems, the MCMC method may need more than 10^6 expensive runs of reservoir simulation and is not practically feasible. Our solution is to employ the reduced terms polynomial chaos proxy introduced by Bazargan et al. [1] to efficiently estimate the posterior probability density function under each cluster and calculate the posterior probability of each model $P(\mathcal{M}_j|\mathcal{D})$.

This work is structured as follows; first we review the mixture modeling approach for clustering, followed by an introductory note on the Bayesian model selection. Then we propose the cluster based polynomial chaos framework for the Bayesian inference. Through examples we will show that the cluster-based polynomial chaos proxy is more efficient in approximating the mixture posterior distribution compared to the standard non-cluster based polynomial chaos proxy. Finally, we apply the cluster-based polynomial chaos proxy to calculate the posterior probability of three different scenarios for the second layer of the Stanford VI reservoir.

2. Mixture Modeling

Mixture modeling is a probabilistic approach to represent the presence of clusters with different probability distribution within an overall distribution of data [25, 26]. Hence, the clusters for a $p-$dimensional dataset with N observations are considered to have different populations while data in each cluster has the same statistics. Accordingly, the data are assumed to be samples from a mixture of K underlying populations, each corresponding to a cluster with a specific probability density function. Hence, the clustering analysis is to determine

- the probability density function in each cluster,
- the number of clusters,
- an optimization algorithm, and

- criteria to select the optimal method [27, 28].

Since there are various optimization methods to the parameter estimation problem as well as numerous models to estimate the probability density function of each cluster, there is a significant amount of opportunity available in the development of the clustering method [27, 28, 25, 26].

In this work, we assume the number of clusters is already determined either by the geological information or the statistical tools for the pattern recognition. Also, the prior distribution of the K underlying populations is considered to be multivariate Gaussian distribution with different statistics. The criteria to select the optimal method is the Bayesian model selection algorithm and the probability density function in each cluster is updated with the posterior probabilities given the observed data $\mathcal{D}$.

3. Bayesian Model Selection

The Bayesian model selection is an alternative approach to the classical *hypothesis testing* [29]. Contrary to the classical likelihood-ratio test, the selection criteria in the Bayesian inference does not solely depend on any single set of parameters. It integrates over all possible parameters in each scenario with respect to their corresponding priors and gives the measure of how well a model is supported by observations.

Assume the realizations from the uncertain domain $\mathbf{z}$ is clustered into K different populations $(\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_K)$, where under each model class $f(\mathbf{z}|\mathcal{M}_i)$ represents the probability distribution of the uncertain parameters. The probability distribution $f(\mathbf{z})$ can be written as the mixture probability density function of the underlying clusters;

$$f(\mathbf{z}) = \sum_{i=1}^{K} P(\mathcal{M}_i) f(\mathbf{z}|\mathcal{M}_i). \tag{2}$$

Let $P(\mathcal{D}|\mathcal{M}_i)$ be the probability distribution (likelihood) of the observed data under model class $\mathcal{M}_i$ and $P(\mathcal{M}_i)$ be the prior probability of model $\mathcal{M}_i$. The fundamental strategy in Bayesian model selection rest on analyzing the posterior probability distribution of models $P(\mathcal{M}_i|\mathcal{D})$. The posterior probability $P(\mathcal{M}_i|\mathcal{D})$ of model $\mathcal{M}_i$ given observed data $\mathcal{D}$ is given by Bayes's theorem:

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{M}_i)P(\mathcal{D}|\mathcal{M}_i)}{\sum_K P(\mathcal{M}_i)P(\mathcal{D}|\mathcal{M}_i)}, \tag{3}$$

where

$$P(\mathcal{D}|\mathcal{M}_i) = \int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) f(\mathbf{z}|\mathcal{M}_i) d\mathbf{z} \tag{4}$$

is the marginal likelihood of $\mathcal{M}_i$. On the basis of observed data $\mathcal{D}$, the plausibility of two different model classes $\mathcal{M}_1$ and $\mathcal{M}_2$ is given by the Bayes factor:

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \tag{5}$$

The Bayes factor updates the prior odds $\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}$ to achieve the posterior odds $\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})}$. By replacing $P(\mathcal{D}|\mathcal{M}_i)$ with the marginal likelihood of equation (4), we obtain

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_1)f(\mathbf{z}|\mathcal{M}_1)d\mathbf{z}}{\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_2)f(\mathbf{z}|\mathcal{M}_2)d\mathbf{z}} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \tag{6}$$

While Bayesian model selection is straightforward in principle, the calculation of the Bayes factor is not. Two challenges in the practical implementation of the Bayesian model selection is the calculation of the marginal likelihood of the equation (4) and the specification of priors. For the selection of priors one simple and popular choice is the uniform prior, where

$$P(\mathcal{M}_1) = P(\mathcal{M}_2) = \ldots = P(\mathcal{M}_K) = \frac{1}{K}. \tag{7}$$

The calculation of the marginal likelihood ($\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_1)f(\mathbf{z}|\mathcal{M}_1)d\mathbf{z}$) often involves multi-dimensional integrals that are hard to evaluate and pose a major challenge to the Bayesian model selection. When the exact calculation of the marginal likelihood is not feasible, the Markov Chain Monte Carlo methods are commonly used to estimate multi-dimensional integrals and evaluate the Bayes factor [30, 31, 32, 24]. Several other schemes have also been developed by Bayesian authors for the numerical calculation of the Bayes factor. In particular, the reversible jump approach [20, 33] and path sampling [34] have gained popularity in recent years.

In high-dimensional problems, one popular approximation for the $P(\mathcal{D}|\mathcal{M}_j)$, when $h(\mathbf{z}) = \log\big(f(\mathcal{D}|\mathbf{z},\mathcal{M}_j)f(\mathbf{z}|\mathcal{M}_j)\big)$ is sufficiently smooth and well-behaved, is obtained by The Laplace method as

$$P(\mathcal{D}|\mathcal{M}_j) \approx (2\pi)^{\frac{-d}{2}}|H(\tilde{\mathbf{z}})|^{\frac{1}{2}}f(\mathcal{D}|\tilde{\mathbf{z}},\mathcal{M}_j)f(\tilde{\mathbf{z}}|\mathcal{M}_j) \tag{8}$$

where d is the dimension of uncertain domain $\mathbf{z}$, $\tilde{\mathbf{z}}$ is the maximum likelihood of $h(\mathbf{z})$ and $H(\tilde{\mathbf{z}})$ is minus the inverse Hessian of $h(\mathbf{z})$ evaluated at the maximum likelihood point [35, 36]. The approximation is obtained by replacing the Taylor series around the maximum likelihood point $h(\mathbf{z}) = h(\tilde{\mathbf{z}}) - \frac{1}{2}(\mathbf{z}-\tilde{\mathbf{z}})^T H(\tilde{\mathbf{z}})(\mathbf{z}-\tilde{\mathbf{z}})$ for $h(\mathbf{z})$ in $P(\mathcal{D}|\mathcal{M}_j) = \int e^{h(\mathbf{z})}d\mathbf{h}(\mathbf{z})$. However, finding the maximum likelihood point $\tilde{\mathbf{z}}$ usually requires expensive numerical computations [35]. Accordingly, further approximations of $P(\mathcal{D}|\mathcal{M}_j)$ can be achieved by replacing $\tilde{\mathbf{z}}$ with $\mathbf{z}^*$ the maximum likelihood estimate of $h^*(\mathbf{z}) = \log\big(f(\mathcal{D}|\mathbf{z},\mathcal{M}_j)\big)$, and $H(\tilde{\mathbf{z}})$ with $H^*(\mathbf{z}^*)$, minus the inverse Hessian of the log likelihood or Fisher's information matrix [35, 36]. Schwarz gave the BIC approximation for (8) when the number of observations are large [14];

$$\log\big(P(\mathcal{D}|\mathcal{M}_j)\big) \approx \log\big(f(\mathcal{D}|\mathbf{z}^*,\mathcal{M}_j)\big) - \frac{d}{2}\log(N) \tag{9}$$

where N is the number of observations. Raftery et al. [59] demonstrated the successful demonstration of Schwarz's approximation in a survival analysis problem. However, McCulloch et al. [37] showed that the BIC approximation may result in poor estimation when the number of observations are small. Since there is explicitly no prior probability distribution

in Schwarz's approximation, it is formally not a Bayesian selection, but it may be implicitly considered as a Bayesian under a "unit information prior" [36, 21] or a "normalized Jeffreys prior" [36, 38].

Several variants of The Laplace approximation have been proposed in the Bayesian literature using linear regression of $h(\mathbf{z})$. In particular, shrinkage estimators (e.g. ridge regression) are very popular [39]. In the family of shrinkage estimators, the Lasso method [40] is the most prominent algorithm which minimizes the least square error of the regression with an upper bound on the uncertain parameters. However, for the reason that the Laplace's approximation is basically a linear estimation around the maximum likelihood point, these methods are not accurate for the cases where $h(\mathbf{z})$ is considerably nonlinear.

For nonlinear models, the efficiency of MCMC to explore the posterior distribution is greatly enhanced when rapidly computable closed form expressions for the marginal likelihoods $f(\mathcal{D}|\mathbf{z}, \mathcal{M}_j)$ are available. In this work, as it will be thoroughly explained in the next section, we utilize the polynomial chaos approximation to achieve an analytical expression for the likelihood and compute the multidimensional integral of the equation (4) by substituting the polynomial chaos proxy for reservoir simulator. As we assume under each model class the prior distributions $P(\mathbf{z}|\mathcal{M}_i)$ are all multivariate Gaussian distributions, the polynomial chaos expansion gives an exponential convergence rate for the estimation of likelihood distribution.

If the object of Bayesian model selection is not only to identify the most probable model but to predict future observations as well, it is best to consider the panoply of models and the inferences or predictions they would give [41, 59]. A formal Bayesian solution to the predictive Bayesian model selection, was first proposed by Leamer [42]. The posterior probability of the mixture distribution $f(\mathbf{z}|\mathcal{D})$ can be expressed as

$$f(\mathbf{z}|\mathcal{D}) = \sum_{i=1}^{K} f(\mathbf{z}|\mathcal{D}, \mathcal{M}_i) P(\mathcal{M}_i|\mathcal{D}). \tag{10}$$

Consequently, if Δ is the quantity of interest, the posterior mean of Δ is given by

$$E(\Delta|\mathcal{D}) = \sum_{i=1}^{K} E(\Delta|\mathcal{D}, \mathcal{M}_i) P(\mathcal{M}_i|\mathcal{D}). \tag{11}$$

Madigan and Raftery [30] demonstrated that averaging over all the models in this fashion provides more accurate predictions. Raftery, Madigan and Hoeting [59] called this method the Bayesian Model Averaging (BMA). Several authors have studied the advantages of BMA and the costs of ignoring model uncertainty [43, 41, 59, 44, 21].

By applying the Baye's rule, we can expand (10) as

$$
\begin{aligned}
f(\mathbf{z}|\mathcal{D}) &= \frac{f(\mathbf{z})}{P(\mathcal{D})} \cdot f(\mathcal{D}|\mathbf{z}) = \frac{f(\mathbf{z})}{P(\mathcal{D})} \Big(\sum_{i=1}^{K} f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) f(\mathcal{M}_i|\mathbf{z}) \Big) \\
&= \frac{f(\mathbf{z})}{f(\mathcal{D})} \Big(\sum_{i=1}^{K} f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) \frac{f(\mathbf{z}|\mathcal{M}_i) P(\mathcal{M}_i)}{f(\mathbf{z})} \Big) \\
&= \sum_{i=1}^{K} f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) \frac{f(\mathbf{z}|\mathcal{M}_i) P(\mathcal{M}_i)}{P(\mathcal{D})}.
\end{aligned}
\tag{12}
$$

where $f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i)$ is the likelihood of $\mathcal{D}$ under each cluster $\mathcal{M}_i$.

4. Bayesian Inference Using The Cluster-based Polynomial Chaos Proxy

Bazargan et al. [1] proposed an efficient method of uncertainty quantification for complex geological structures using the polynomial chaos expansion. The original polynomial chaos expansion was first proposed by Wiener [60] to represent a general second-order random variable $y = \mathbf{G}(\theta)$, viewed as a function of random variable θ, in the following form:

$$
\begin{aligned}
y = a_0 H_0 &+ \sum_{i_1=1}^{\infty} a_{i_1} H_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1,i_2} H_2(\xi_{i_1}, \xi_{i_2}) \\
&+ \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1,i_2,i_3} H_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \dots .
\end{aligned}
\tag{13}
$$

where $H_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ denotes orthogonal Hermite polynomials of order n and $(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ are multi-dimensional independent Gaussian random variables with zero mean and unit variance. The Hermite polynomials of order n, $H_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ can be derived by the following formulation [60]:

$$
H_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n}) = e^{(\frac{1}{2}\xi^T \xi)} (-1)^n \frac{\partial^n}{\partial \xi_{i_1} \partial \xi_{i_2} \dots \partial \xi_{i_n}} e^{-(\frac{1}{2}\xi^T \xi)}.
\tag{14}
$$

Here, ξ denotes the vector of n Gaussian random variables. Under the Gaussian probability measure $(w(\xi) : \mathcal{N}(0, I))$, Hermite polynomials are orthogonal to each other and form a complete basis of the Hilbert space. The orthogonality of Hermite polynomials under the Gaussian measure implies:

$$
\langle H_i(\xi), H_j(\xi) \rangle = \int_{-\infty}^{\infty} H_i(\xi) H_j(\xi) e^{-\frac{\xi^2}{2\sigma^2}} d\xi = 0 \quad i \neq j,
\tag{15}
$$

and also,

$$
\langle H_i(\xi), H_i(\xi) \rangle = \int_{-\infty}^{\infty} H_i(\xi) H_i(\xi) \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2\sigma^2}} d\xi = i! \quad \forall i \geq 1.
\tag{16}
$$

Since the polynomial bases are all orthogonal to each other, the deterministic coefficients of (13) can be computed using a Galerkin projection scheme:

$$
\begin{aligned}
a_{i_1,i_2,\ldots,i_n} &= \frac{\langle H_n, y\rangle}{\langle H_n, H_n\rangle} = \frac{\int_{-\infty}^{\infty} H_n(\xi) y w(\xi) d\xi}{\int_{-\infty}^{\infty} H_n^2(\xi) w(\xi) d\xi} \\
&= \frac{\mathbf{E}[H_n(\xi) y]}{\mathbf{E}[H_n^2(\xi)]} \simeq \frac{\sum_{i=1}^{N_s} H_n(\xi_i) y_i}{\sum_{i=1}^{N_s} H_n^2(\xi_i)}.
\end{aligned}
\tag{17}
$$

However, for computing the coefficients correctly, the convergence of the right hand side of the equation has to be studied. For higher order Hermite polynomials, large number of trial runs may be required to reach the convergence. For a given number of simulation runs, the coefficients associated with the higher order of PCE terms tend to have larger estimation errors compared with the lower order terms, and therefore, there is a balance between the number of training runs and the maximum possible order of PCE that can be reliably used. To resolve this issue, various methods have been proposed in the literature. Bazargan et al. [1] found an upper bound for the polynomial chaos coefficients and based on that they proposed an impact factor to drop the irrelevant terms. They demonstrated that the method works efficiently when the polynomial basis is adapted to the input probability distribution that makes the polynomial chaos representation highly sparse.

However, for practical problems we are provided with a set of realizations rather than a specific probability distribution and we need to use numerical methods to estimate the density and construct polynomial chaos basis adaptively in a way that the polynomial chaos representation becomes fairly sparse. One special case is when the realizations can be distinctly clustered into K different clusters where under each cluster the probability distribution can be reasonably approximated by Gaussian distribution. In theory, all probability density functions can be expressed in terms of weighted Gaussian probability density functions, centered at each of the data points and then taking the average to yield Gaussian kernel density estimation [45]. However, the success of Gaussian kernel density estimation relies heavily on the appropriate selection of the bandwidth parameter which is known to be extremely difficult for heavy-tailed distributions or distributions that have discontinuities in their density functions. Accordingly, the model-based clustering approach can be extended to problems where the number of clusters K is unknown and can be optimized by studying the distribution of each cluster. In this work, without loss of generality, we consider geological structures where the number of clusters K is pre-determined by geological intuition.

The objective is to compare different models (clusters) by the Bayesian geological model selection discussed in Section 3.

To estimate the posterior probability of each cluster, we use the reduced terms polynomial chaos expansion with Hermite polynomials as a proxy substitute of the actual reservoir simulator. Having an analytical expression for the posterior probability distribution of each cluster, and consequently the mixture, we can readily compute the posterior probability of

each model.

We assume the probability distribution of $\mathbf{z}(\xi)$ under each cluster to be $P(\mathbf{z}|\mathcal{M}_j) : \mathcal{N}(\mu_j, \Sigma_j)$. Based on the given trial runs for each cluster, the polynomial chaos proxy $PC_{\mathcal{M}_j}$ for the cluster $\mathcal{M}_j$ is constructed via regression-based PCM as discussed in Bazargan et al [1]. Regarding the exponential convergence rate of the standard Hermite polynomial chaos for multivariate Gaussian random variables in each cluster $\mathcal{N}(\mu_j, \Sigma_j)$, even low order polynomial chaos suffices to give a reasonably accurate approximation.

We write the common misfit formulation given the observed data $\mathcal{D}$ under each cluster $\mathcal{M}_j$, and replace the actual model with its corresponding polynomial chaos approximation $PC_{\mathcal{M}_j}$ [1];

$$\begin{aligned} S(\mathbf{z}(\xi)|\mathcal{M}_j) &= \sum_i -\frac{(\mathbf{G}_i(\mathbf{z}) - \mathcal{D}_i)^2}{2\sigma_i^2} \simeq \sum_i -\frac{(\mathbf{PC}_{\mathcal{M}_\mathbf{j}}(\mathbf{z}) - \mathcal{D}_i)^2}{2\sigma_i^2} \\ &= S_0^{\mathcal{M}_j} + \sum_{i=1}^{N} S_{i,1}^{\mathcal{M}_j}\xi_i + \sum_{i=1}^{N} S_{ii,2}^{\mathcal{M}_j}(\xi_i^2 - 1) + \sum_{i=1}^{N-1}\sum_{k>i}^{N} S_{ik,2}^{\mathcal{M}_j}\xi_i\xi_k + \dots \end{aligned} \tag{18}$$

The likelihood of the observed data under each cluster can be expressed as

$$f(\mathcal{D}|\mathbf{z}(\xi), \mathcal{M}_j) = ke^{S(\mathbf{z}(\xi)|\mathcal{M}_j)}, \tag{19}$$

where the misfit surface under each cluster $S(\mathbf{z}(\xi)|\mathcal{M}_j)$ can be computed via (18). Therefore, the Bayes factor of the equation (6) can be written as

$$\begin{aligned} \frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} &= \frac{\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_1) f(\mathbf{z}|\mathcal{M}_1) d\mathbf{z}}{\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_2) f(\mathbf{z}|\mathcal{M}_2) d\mathbf{z}} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)} \\ &\simeq \frac{\int e^{S(\mathbf{z}(\xi)|\mathcal{M}_1)} f(\mathbf{z}(\xi)|\mathcal{M}_1) d\mathbf{z}}{\int e^{S(\mathbf{z}(\xi)|\mathcal{M}_2)} f(\mathbf{z}(\xi)|\mathcal{M}_2) d\mathbf{z}} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \end{aligned} \tag{20}$$

Since we obtained an analytical expression for $S(\mathbf{z}(\xi)|\mathcal{M}_j)$ and $f(\mathbf{z}(\xi)|\mathcal{M}_j) : \mathcal{N}(\mu_j, \Sigma_j)$, the above integration can be efficiently calculated using the Monte Carlo techniques. Accordingly, we can also achieve an analytical expression for the posterior probability of the mixture distribution by (12):

$$f(\mathbf{z}|\mathcal{D}) \simeq k \sum_{i=1}^{K} e^{S(\mathbf{z}(\xi)|\mathcal{M}_j)} f(\mathbf{z}(\xi)|\mathcal{M}_i) P(\mathcal{M}_i) \tag{21}$$

The above formulation for the posterior mixture distribution is also called "Bayesian Model Averaging" [59].

Example 4.1. ***Bayesian model selection for the mixture distribution of two Gaussian***

Let z be a random variable whose probability distribution function can be expressed as a mixture of two Gaussian probability distribution;

$$f(z) = \frac{1}{2}\Big(\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-2)^2}{2}}\Big) + \frac{1}{2}\Big(\frac{1}{\sqrt{2\pi}}e^{-\frac{(z+2)^2}{2}}\Big). \tag{22}$$

Consequently, z can be effectively clustered into two different model, $\mathcal{M}_1$ and $\mathcal{M}_2$ where $P(\mathcal{M}_1) = P(\mathcal{M}_2) = \frac{1}{2}$ and

$$\begin{aligned} f(z|\mathcal{M}_1) &: \mathcal{N}(2,1) \\ f(z|\mathcal{M}_2) &: \mathcal{N}(-2,1). \end{aligned} \tag{23}$$

Assume the underlying physical system is $G(z) = e^z$ and the observed data $\mathcal{D} = 2$. For the likelihood distribution we consider the simple misfit formulation [1]:

$$f(\mathcal{D}|z) = ke^{-\frac{(G(z)-D)^2}{2\sigma^2}} = ke^{-\frac{(e^z-D)^2}{2\sigma^2}}. \tag{24}$$

Let the variance of the error in observation be $\sigma^2 = 0.5$. We want to calculate the pos-

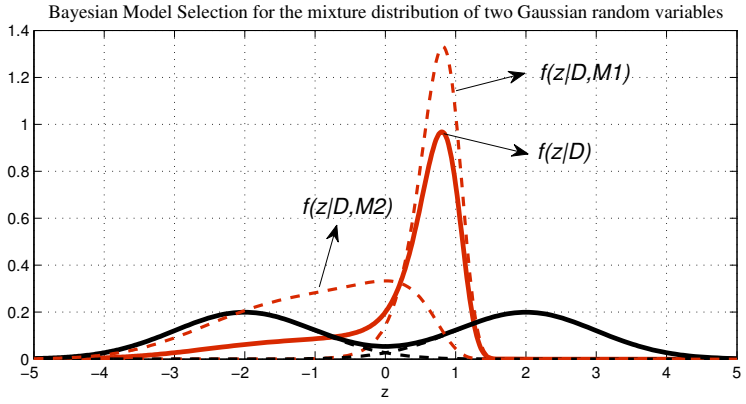

Figure 1: The red line shows the posterior probability of the mixture distribution of Example 4.1. The posterior probability under each cluster is shown with the red-dotted lines.

terior probability of each cluster, Bayes factor and the posterior probability of the mixture distribution given the observed data $\mathcal{D}$. From the equation (6) we obtain

$$\begin{aligned} \frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} &= \frac{\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_1)f(\mathbf{z}|\mathcal{M}_1)dz}{\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_2)f(\mathbf{z}|\mathcal{M}_2)dz} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)} \\ &= \frac{\int e^{-\frac{(e^z-2)^2}{2\sigma^2}}e^{-\frac{(z-2)^2}{2}}d\mathbf{z}}{\int e^{-\frac{(e^z-2)^2}{2\sigma^2}}e^{-\frac{(z+2)^2}{2}}d\mathbf{z}} = 2.327. \end{aligned} \tag{25}$$

We also know the fact that $P(\mathcal{M}_1|\mathcal{D}) + P(\mathcal{M}_2|\mathcal{D}) = 1$, then we achieve

$$\begin{aligned} P(\mathcal{M}_1|\mathcal{D}) &\simeq 0.6994 \\ P(\mathcal{M}_2|\mathcal{D}) &\simeq 0.3006. \end{aligned} \tag{26}$$

The mixture posterior probability distribution of z can also be computed via (12)

$$f(z|\mathcal{D}) = const \cdot \big(\frac{1}{2} \cdot e^{-\frac{(e^z-2)^2}{2\sigma^2}} e^{-\frac{(z-2)^2}{2}} + \frac{1}{2} \cdot e^{-\frac{(e^z-2)^2}{2\sigma^2}} e^{-\frac{(z+2)^2}{2}}\big). \tag{27}$$

Figure 1 illustrates the prior and the posterior probability of z under each cluster, in addition to the prior and the posterior probability of the mixture distribution.

Under each cluster $\mathbf{z}$ is Gaussian and can be expressed as:

$$\mathbf{z} = \begin{cases} \xi + 2 & \mathbf{z} \in \mathcal{M}_1 \\ \xi - 2 & \mathbf{z} \in \mathcal{M}_2 \end{cases} \tag{28}$$

where ξ is a normal random variable $\mathcal{N}(0,1)$. By the equation 17, the standard polynomial chaos expansion of $G(\xi) = e^{\xi}$ can be obtained as

$$PC(G(\xi)) = \sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi) = e^{\frac{1}{2}}(1 + \xi + \frac{(\xi^2-1)}{2!} + \frac{(\xi^3-3\xi)}{3!} + \frac{(\xi^4-6\xi^2+3)}{4!} + \ldots). \tag{29}$$

Hence the polynomial chaos proxy under each cluster can be deduced via (29)

$$PC(G(\mathbf{z})) = \begin{cases} e^2\Big(\sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi)\Big) & \mathbf{z} \in \mathcal{M}_1 \\ e^{-2}\Big(\sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi)\Big) & \mathbf{z} \in \mathcal{M}_2 \end{cases} \tag{30}$$

Accordingly, the polynomial chaos proxy under each cluster can be expanded as

$$PC(G(\mathbf{z})) = \begin{cases} e^{\frac{5}{2}}\Big(1 + (z-2) + \frac{\big((z-2)^2-1\big)}{2!} + \frac{\big((z-2)^3-3(z-2)\big)}{3!} + \ldots\Big) & \mathbf{z} \in \mathcal{M}_1 \\ e^{\frac{-3}{2}}\Big(1 + (z+2) + \frac{\big((z+2)^2-1\big)}{2!} + \frac{\big((z+2)^3-3(z+2)\big)}{3!} + \ldots\Big) & \mathbf{z} \in \mathcal{M}_2 \end{cases} \tag{31}$$

The likelihood under each cluster can be approximated by replacing $G(\mathbf{z})$ of the equation (24) with its corresponding polynomial chaos proxy,

$$\begin{aligned} f(\mathcal{D}|\mathbf{z},\mathcal{M}_1) &\simeq k \cdot e^{-\frac{(PC_{\mathcal{M}_1}(z)-D)^2}{2\sigma^2}} = k \cdot e^{-\frac{(e^{\frac{5}{2}}(1+(z-2)+\frac{((z-2)^2-1)}{2!}+\ldots-D)^2}{2\sigma^2}} \\ f(\mathcal{D}|\mathbf{z},\mathcal{M}_2) &\simeq k \cdot e^{-\frac{(PC_{\mathcal{M}_2}(z)-D)^2}{2\sigma^2}} = k \cdot e^{-\frac{(e^{\frac{-3}{2}}(1+(z+2)+\frac{((z+2)^2-1)}{2!}+\ldots-D)^2}{2\sigma^2}} \end{aligned} \tag{32}$$

Under each cluster, the Cameron-Martin Theorem [61] implies the exponential convergence rate for the polynomial chaos expansion to $G(\mathbf{z})$. Consequently, the posterior distribution under each cluster can be obtained by equation (12). Figure 2 demonstrates the fast convergence rate of the estimation of the posterior distribution under each cluster using the polynomial chaos proxy to the exact answer derived in Example 4.1. The estimation of the

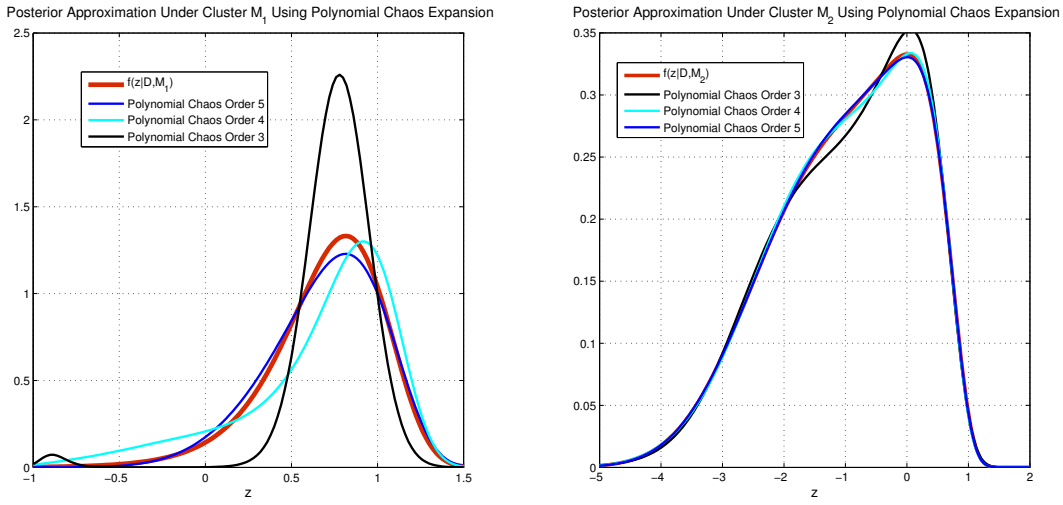

Figure 2: The posterior estimation under cluster $\mathcal{M}_1$ and $\mathcal{M}_2$ using the polynomial chaos expansion.

posterior probability of the mixture distribution using the cluster-based polynomial chaos proxy is shown on Figure 3.

However, in the non-cluster approach, the convergence rate of the standard polynomial chaos expansion is slow. Figure 4 illustrates the slow convergence of non-cluster based polynomial chaos proxy. The standard polynomial chaos expansion is the optimal choice for Gaussian random variables but for non-Gaussian random variables (e.g., the mixture of Gaussian) a higher order of polynomial chaos expansion may be required for the accurate approximation of the target distribution.

Figure 5 compares the accuracy in the estimation of the posterior probability of the mixture distribution using the cluster-based polynomial chaos proxy versus the standard non-cluster polynomial chaos proxy. It demonstrates that even the low order (4^{th} order) of the polynomial chaos expansion in the cluster-based approach gives a better approximation of posterior distribution than the polynomial chaos expansion of 12^{th} order in the non-cluster approach.

Accordingly, we can approximate the Bayes factor of the equation (25) by replacing $G(\mathbf{z})$ with the polynomial chaos proxy under each cluster. Table 1 compares the Bayes factor approximation by the polynomial chaos proxy of different order to the exact answer computed in Example 4.1. It is also instructive to compute the Bayes factor by the Laplace approximation and compare the accuracy with the polynomial chaos approximation. The Laplace approximation under each cluster can be obtained by first computing the log of likelihood $h(\mathbf{z}|\mathcal{M}_i) = \log(f(\mathcal{D}|\mathbf{z},\mathcal{M}_i)f(\mathbf{z}|\mathcal{M}_i))$ for each cluster;

$$
\begin{aligned}
h(\mathbf{z}|\mathcal{M}_1) &= \log(e^{-\frac{(e^{\mathbf{z}}-2)^2}{2\sigma^2}} e^{-\frac{(\mathbf{z}-2)^2}{2}}) = -\frac{(\mathbf{z}-2)^2}{2} - \frac{(e^{\mathbf{z}}-2)^2}{2\sigma^2} \\
h(\mathbf{z}|\mathcal{M}_2) &= \log(e^{-\frac{(e^{\mathbf{z}}-2)^2}{2\sigma^2}} e^{-\frac{(\mathbf{z}+2)^2}{2}}) = -\frac{(\mathbf{z}+2)^2}{2} - \frac{(e^{\mathbf{z}}-2)^2}{2\sigma^2}.
\end{aligned}
\tag{33}
$$

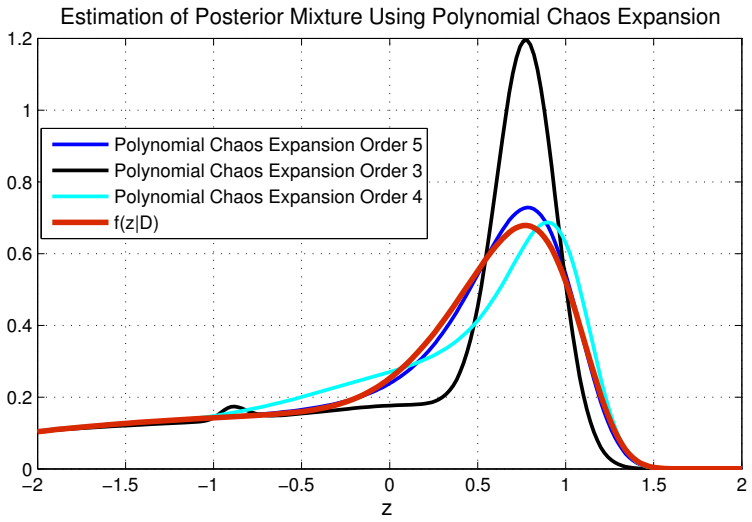

Figure 3: The estimation of the posterior probability of the mixture distribution using the cluster-based polynomial chaos proxy.

Next we find the maximum likelihood point under each cluster by solving $h'(\mathbf{z}|\mathcal{M}_i) = 0$ *where* $\sigma^2 = 0.5$*;*

$$\begin{aligned} h'(\mathbf{z}_1^*|\mathcal{M}_1) &= -(\mathbf{z}_1^* - 2) - \frac{(e^{\mathbf{z}_1^*} - 2)e^{\mathbf{z}_1^*}}{\sigma^2} = 0 \Rightarrow \mathbf{z}_1^* = 0.8161 \\ h'(\mathbf{z}_2^*|\mathcal{M}_2) &= -(\mathbf{z}_2^* + 2) - \frac{(e^{\mathbf{z}_2^*} - 2)e^{\mathbf{z}_2^*}}{\sigma^2} = 0 \Rightarrow \mathbf{z}_2^* = 0. \end{aligned} \tag{34}$$

The Laplace method estimates $h(\mathbf{z}|\mathcal{M}_i)$ *by the second order Taylor expansion around the maximum likelihood point;* $h(\mathbf{z}|\mathcal{M}_i) \simeq h(\mathbf{z}^*|\mathcal{M}_i) + \frac{1}{2}h''(\mathbf{z}^*|\mathcal{M}_i)(\mathbf{z}-\mathbf{z}^*)^2$*, noting that* $h'(\mathbf{z}^*|\mathcal{M}_i) = 0$*;*

$$\begin{aligned} h(\mathbf{z}|\mathcal{M}_1) &\simeq -0.76927 - \frac{12.413}{2}(\mathbf{z} - 0.8161)^2 \\ h(\mathbf{z}|\mathcal{M}_2) &\simeq -3 - \frac{1}{2}\mathbf{z}^2. \end{aligned} \tag{35}$$

Consequently, we can approximately compute the Bayes factor via The Laplace method as

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} \simeq \frac{\int e^{h(\mathbf{z}|\mathcal{M}_1)}d\mathbf{z}}{\int e^{h(\mathbf{z}|\mathcal{M}_2)}d\mathbf{z}} = \frac{\int e^{-0.76927-\frac{12.413}{2}(\mathbf{z}-0.8161)^2}d\mathbf{z}}{\int e^{-3-\frac{1}{2}\mathbf{z}^2}d\mathbf{z}} \simeq 2.671. \tag{36}$$

Table 1 shows that the polynomial chaos approximation is generally more accurate than The Laplace method to compute the Bayes factor. This is mainly due to the fact that in approximating $G(\mathbf{z})$ *the polynomial chaos proxy takes the nonlinear terms into account while the Laplace method uses a linear estimation around the maximum likelihood point.*

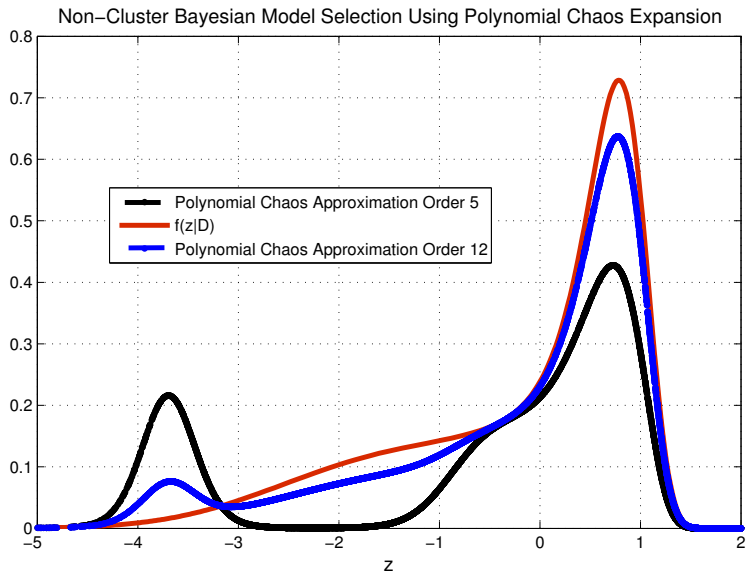

Figure 4: The estimation of the posterior probability of the mixture distribution using the standard non-cluster based polynomial chaos expansion.

Table 1: Bayes factor estimation using the polynomial chaos proxy compared to The Laplace method

	The Bayes Factor
The exact solution	2.327
The Laplace method	2.671
Polynomial chaos proxy order 3	2.494
Polynomial chaos proxy order 4	2.383
Polynomial chaos proxy order 5	2.327

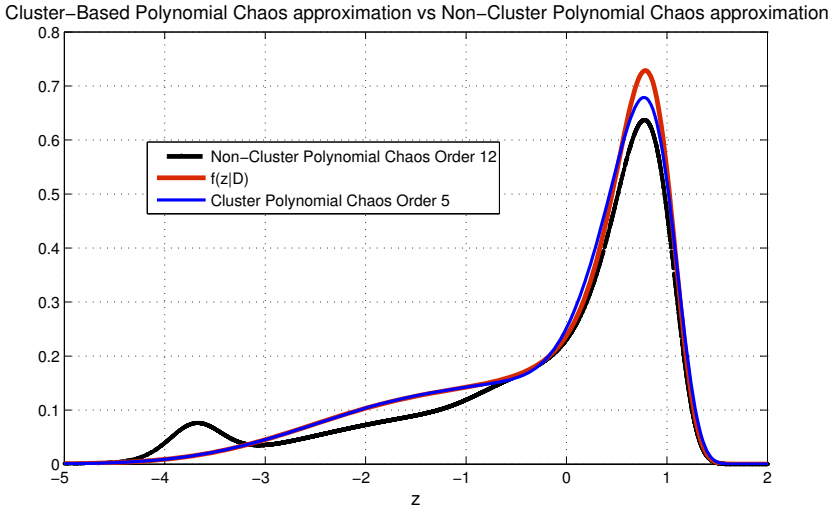

Figure 5: The cluster-based polynomial chaos proxy is more efficient that the non-cluster approach using the standard polynomial chaos expansion.

5. Case Study

Different interpretations of the sedimentary environments are one of the largest sources of uncertainty in reservoir modeling. Several plausible scenarios may be offered as reservoir interpretations based on seismic data, well log, etc. In the context of facies modeling, this could result in several descriptions of facies structures, associations, connectivity and geometry. One quantitative approach to handle different geological scenarios is to provide alternative training images from which several reservoir models can be generated. The problem we address here is to quantitatively compare the plausibility of these training images and also sample from the posterior distribution within each scenario.

The pioneering work to integrate multiple reservoir descriptions using alternative training images was proposed by Caers and Scheidt (2008) [46]. They used a distance metric approach based on multi-dimensional scaling (MDS) to explore the space of the reservoir realizations. The method was further developed in [47] where MDS was used to compare different production responses of realizations obtained from different scenarios with the actual production data. Rojas [48] used a multi-class Support Vector Machine Classifier (SVM) [49] to differentiate between facies scenarios represented by multiple training images. Lately, Park et al. [50] proposed a Bayesian-based approach to calculate the posterior possibility of each scenario using the probability perturbation method. Here we use the cluster-based polynomial chaos proxy for Bayesian model selection. The case study demonstrates how the approach helps to quantify uncertainty between different training images.

We study the plausibility of three training images based on different geological interpretation of the second layer of synthetic Stanford VI reservoir ([51]). The reservoir is $3.75Km$ wide (East-West) and $5.0Km$ (North-South) long, with a shallowest top depth of 2.5 Km and deepest top depth of $2.7Km$. It consists of three layers with thickness of $80m$, $40m$ and $80m$. The Stanford VI reservoir is discretized into $150 \times 200 \times 200$ cells while the dimension of the grid cell is $25m$ in the x and y direction and $1m$ in the z direction. The stratigraphy of the Stanford VI reservoir shows a prograding fluvial channel system, where deltaic deposits represented in the third layer were deposited first and followed by meandering channels in the second layer and sinuous channel in the first layer. See [51] for the detailed description of the Stanford VI reservoir.

The second layer consists of meandering channel represented by four facies: the floodplain (shale deposits), the point bar (sand deposits that occur along the convex inner edges of the meanders of channels), the channel (sand deposits), and the boundary (shale deposits). Similar to Rojas [48], a simplified description of the second layer, which entails only three facies; the floodplain, the pointbar and the channel are modeled using three different training images, shown in Figure 7, Figure 8 and Figure 9. We choose equal prior probabilities for each of the three scenarios $P(\mathcal{M}_1) = P(\mathcal{M}_2) = P(\mathcal{M}_3) = \frac{1}{3}$, and calculate the posterior probability of each scenario given the observed data $P(\mathcal{M}_1|\mathcal{D}), P(\mathcal{M}_2|\mathcal{D}), P(\mathcal{M}_3|\mathcal{D})$. The observed data includes the production profile of the second layer of the reservoir for 2000

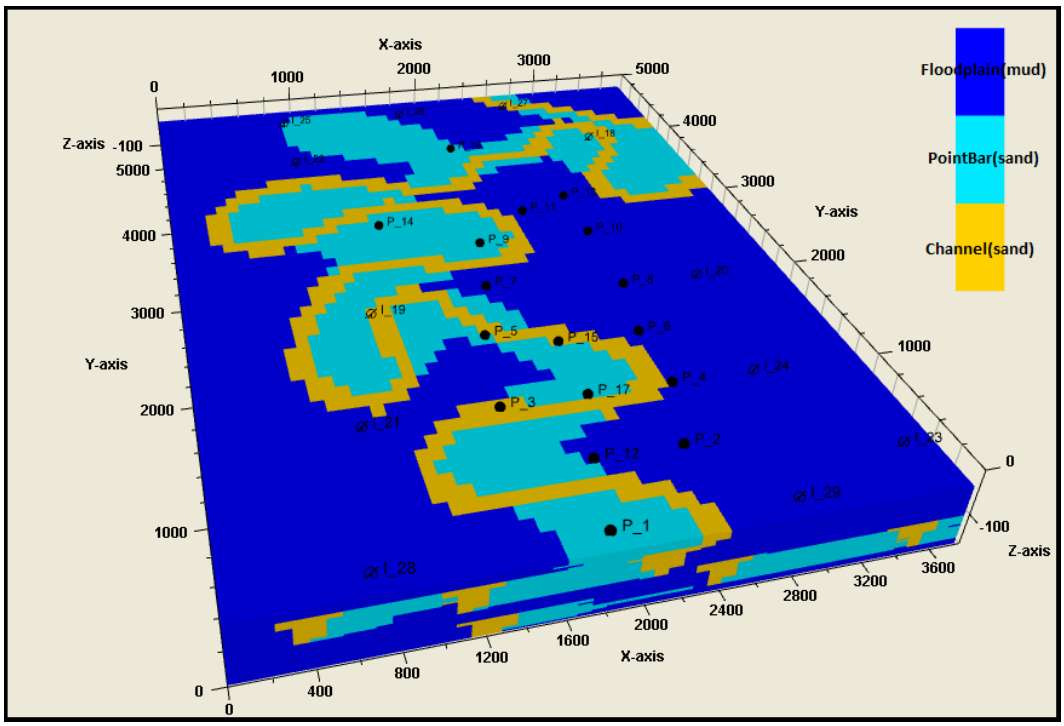

Figure 6: The reference case for the second layer of Stanford VI, with three facies: the floodplain, the point bar and the channel.

days. The truth case used to generate the observation (the oil flow rate, the water cut and the pressure of all wells for 2000 days) is shown in Figure 6.

To calculate the posterior probability of each scenario, we generate $N = 500$ realizations for each training image using *snesim* algorithm [52]. In *snesim* the training image is scanned and the conditional probabilities are constructed based on the conditional proportions from a training image depicting the geometry and distribution of objects deemed to prevail in the actual field.

All the facies realizations are then populated with three petrophysical properties: porosity, density and permeability as described in [51]. Hence, under each scenario we obtain 500 reservoir models. In Figure 11 we apply the principal component analysis to all 1500 facies realizations and plot the first two components. Three distinct clusters can be identified vividly and it demonstrates that the first two eigenvectors are enough to separate the realizations corresponding to the three training images. Figure 12 shows the Scree-graph of the realizations. As the realizations are well separated in three distinct clusters, then the differences between the groups will be picked up by the significant components of PCA.

Here we follow the procedure described in Section 4 which involves the following steps:

- **Dimensionality Reduction**
  Under each cluster we apply the principal component analysis and discard the small eigenvalues. The within-groups covariance matrix describes the average variation of

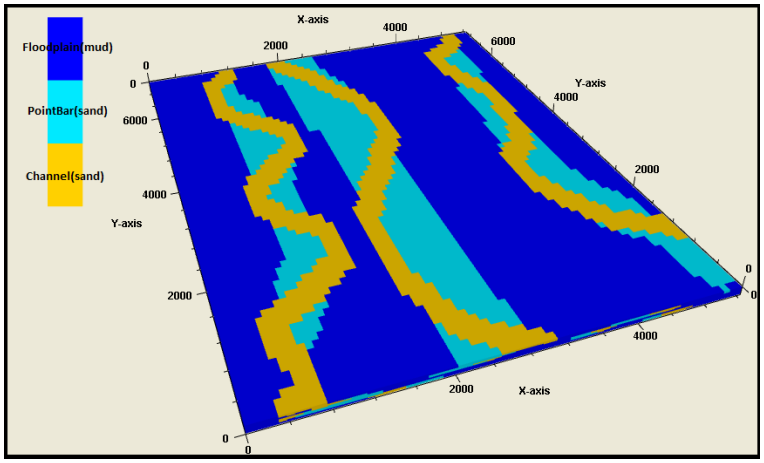

Figure 7: The training image of scenario $\mathcal{M}_1$

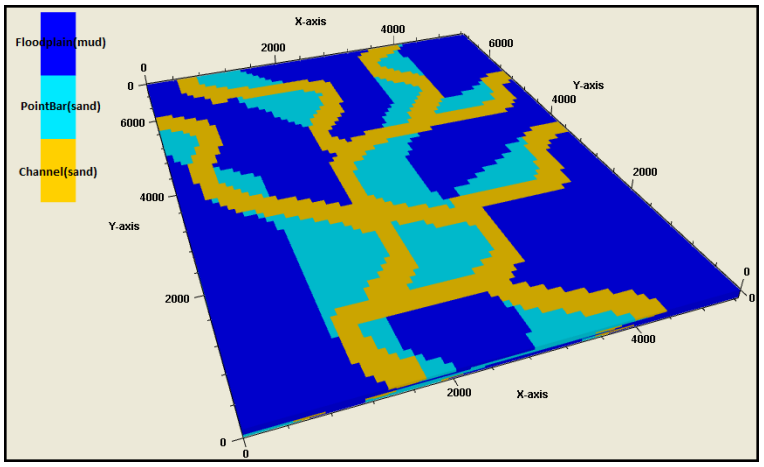

Figure 8: The training image of scenario $\mathcal{M}_2$

each group about its respective group mean. For this specific case, the within-groups covariance matrices of the three training images are similar to each other. Figure 13 illustrates the scree graph of the within-groups covariance matrix. For the Bayesian model selection, we preserve 10 eigenvectors under each cluster as the scree plots essentially flatten out thereafter. However, we compute the Bayes factor while preserving 2, 5, 10, 20 and 40 components.

- **Polynomial Chaos Proxy**
  Using ECLIPSE for the flow simulation, we obtain the production profile of all the wells for the 1500 reservoir models generated earlier (trial runs). Consequently under each cluster, we apply the regression-based PCM using 1500 trial runs to compute the coefficients of the polynomial chaos representation for the oil and water flow rates at each well. For example, for the oil flow rate of the well P_1, we obtain three different

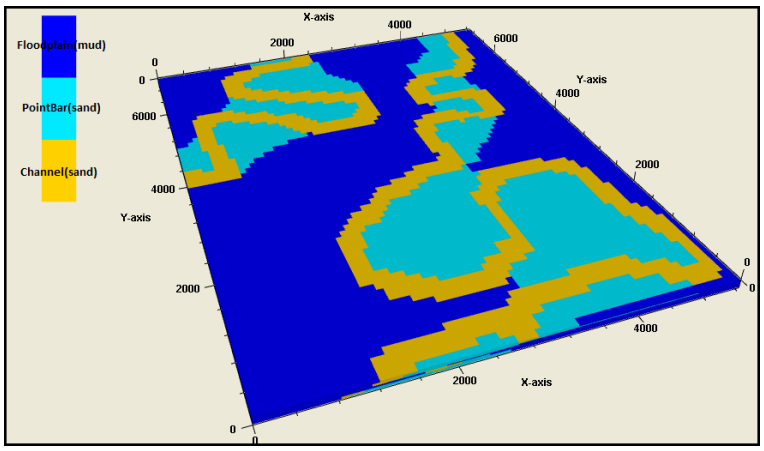

Figure 9: The training image of scenario $\mathcal{M}_3$

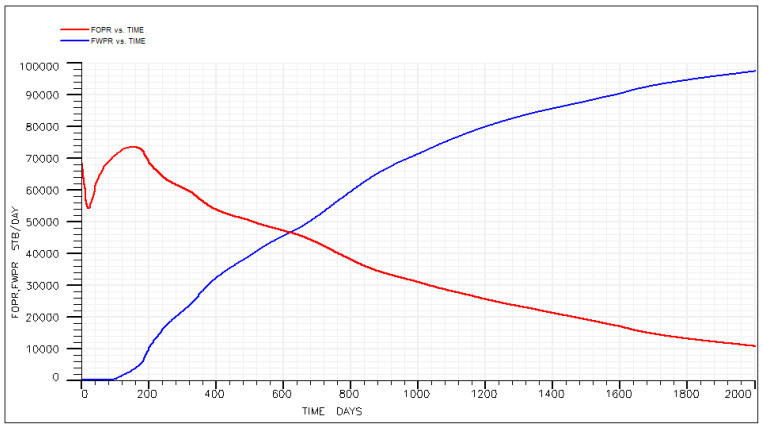

Figure 10: The cumulative oil (red) and water (blue) production rate of the second layer of Stanford VI for 2000 days.

expressions for the polynomial chaos expansion as:

$$\begin{aligned}
\mathbf{Q^{op1}}(\xi,t|\mathcal{M}_1) &= Q_0^{op1|\mathcal{M}_1}(t) + \sum_{i=1}^{P} Q_{i,1}^{op1|\mathcal{M}_1}(t)\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1|\mathcal{M}_1}(t)(\xi_i^2-1) + \dots \\
\mathbf{Q^{op1}}(\xi,t|\mathcal{M}_2) &= Q_0^{op1|\mathcal{M}_2}(t) + \sum_{i=1}^{P} Q_{i,1}^{op1|\mathcal{M}_2}(t)\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1|\mathcal{M}_2}(t)(\xi_i^2-1) + \dots \\
\mathbf{Q^{op1}}(\xi,t|\mathcal{M}_3) &= Q_0^{op1|\mathcal{M}_3}(t) + \sum_{i=1}^{P} Q_{i,1}^{op1|\mathcal{M}_3}(t)\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1|\mathcal{M}_3}(t)(\xi_i^2-1) + \dots
\end{aligned} \tag{37}$$

Since the prior distribution under each cluster is assumed to be multivariate Gaussian, the Cameron-Martin Theorem [61] implies that the rate of convergence of the standard polynomial chaos is exponential. Hence, even a low order polynomial chaos will be reasonably accurate for the Bayes factor calculation. However, we examine the polynomial chaos expansion of order 2, 4, 6 and 8 respectively. For the 8^{th} order polynomial chaos

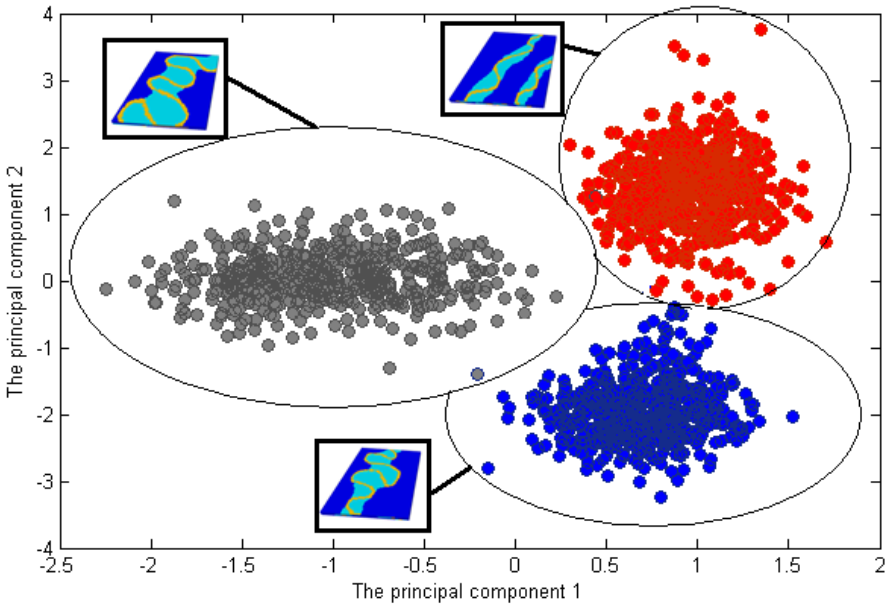

Figure 11: The principal component analysis for the realizations generated from three different training images.

expansion, to reduce the number of the terms, we use the impact factor introduced in [1] to discard the irrelevant terms.

Using the production data of the truth case (Figure 6) and equation (18), we obtain analytical expressions for the misfit surface under each cluster:

$$\begin{aligned}
\mathbf{S}(\xi|\mathcal{M}_\mathbf{1}) &= S_0^{\mathcal{M}_1} + \sum_{i=1}^{P} S_{i,1}^{\mathcal{M}_1}\xi_i + \sum_{i=1}^{P} S_{ii,2}^{\mathcal{M}_1}(\xi_i^2 - 1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} S_{ij,2}^{\mathcal{M}_1}(\xi_i\xi_j) + \dots \\
\mathbf{S}(\xi|\mathcal{M}_\mathbf{2}) &= S_0^{\mathcal{M}_2} + \sum_{i=1}^{P} S_{i,1}^{\mathcal{M}_2}\xi_i + \sum_{i=1}^{P} S_{ii,2}^{\mathcal{M}_2}(\xi_i^2 - 1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} S_{ij,2}^{\mathcal{M}_2}(\xi_i\xi_j) + \dots \\
\mathbf{S}(\xi|\mathcal{M}_\mathbf{3}) &= S_0^{\mathcal{M}_3} + \sum_{i=1}^{P} S_{i,1}^{\mathcal{M}_3}\xi_i + \sum_{i=1}^{P} S_{ii,2}^{\mathcal{M}_3}(\xi_i^2 - 1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} S_{ij,2}^{\mathcal{M}_3}(\xi_i\xi_j) + \dots
\end{aligned} \tag{38}$$

We define the observational error $\sigma^2(t)$ to be multiplicative, i.e. $\sigma^2(t)$ for the oil flow rate of the well P_1 is defined as

$$\sigma(t) = \sigma_D \, Q_{obs}^{op1}(t) \tag{39}$$

where $Q_{obs}^{op1}(t) \in \mathcal{D}$ is the observed oil flow rate of P_1 (shown in Figure 14) and σ_D is the dimensionless observational error and assumed to be a constant for all the wells. We solve the problem for $\sigma_D^2 = 0.1, 0.15, 0.2$ and 0.25.

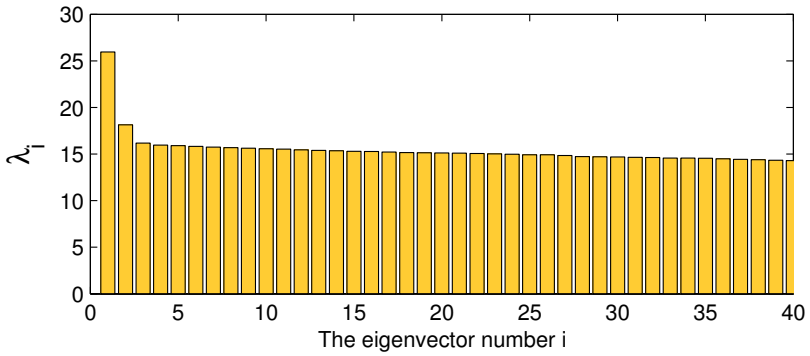

Figure 12: The scree graph of the 1500 realizations generated from three different training images.

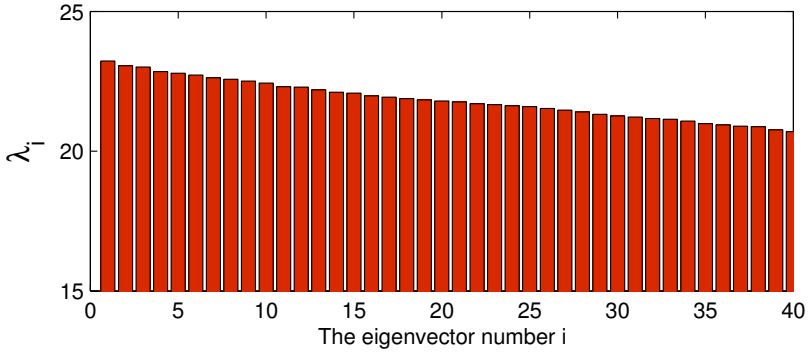

Figure 13: The scree graph of the realizations generated from the training image of scenario $\mathcal{M}_1$.

- **Markov Chain Monte Carlo Sampling**
  Using equation (19), we obtain an analytical expression for the posterior distribution of the reduced dimension parameters ξ under each cluster, from which we can generate samples from the posterior reservoir models by the inversion map. Figure 15 shows samples from the posterior facies model under cluster $\mathcal{M}_1$ while 10 eigenvector is preserved and the order of the polynomial chaos proxy is 6. Figure 16 and 17 illustrates samples from the posterior facies models under cluster $\mathcal{M}_2$ and $\mathcal{M}_3$ respectively.

- **Bayesian Model Selection**
  The high-dimensional integral of the equation (6) can be calculated using the numerical techniques developed in the literature [53, 54, 55, 56, 57]. Here we use the Quasi Monte Carlo [58] technique with the low discrepancy Sobol sequence [53] whose rate of convergence is $\mathcal{O}(\frac{1}{N_s})$. N_s is the number of times that integrand has to be evaluated. Since running the polynomial chaos proxy is cheap, we can efficiently evaluate the integrand more than 10^9 times to achieve the desired accuracy.

  The exact evaluation of the high-dimensional integral of (6) is not practically feasible. To examine the accuracy of the polynomial chaos proxy in the calculation of the Bayes factor, we retain 10 random variables of the reduced dimension space and perform a

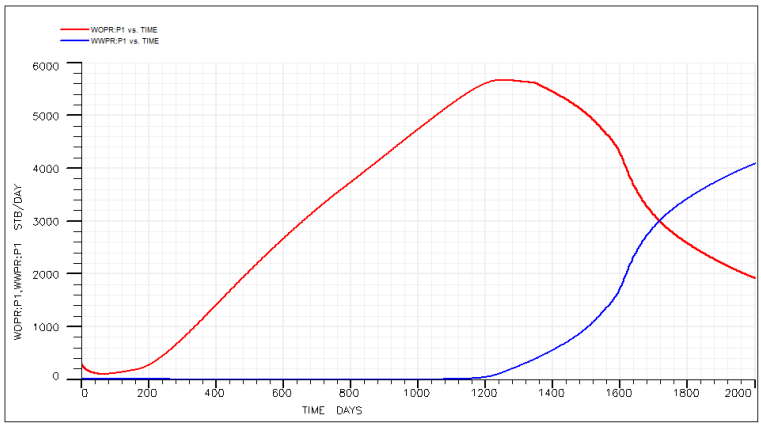

Figure 14: The oil (red) and water (blue) production rate of the well P1 for 2000 days. The observed data includes the production profile (the oil flow rate, the water cut and the pressure) of all the wells for 2000 days.

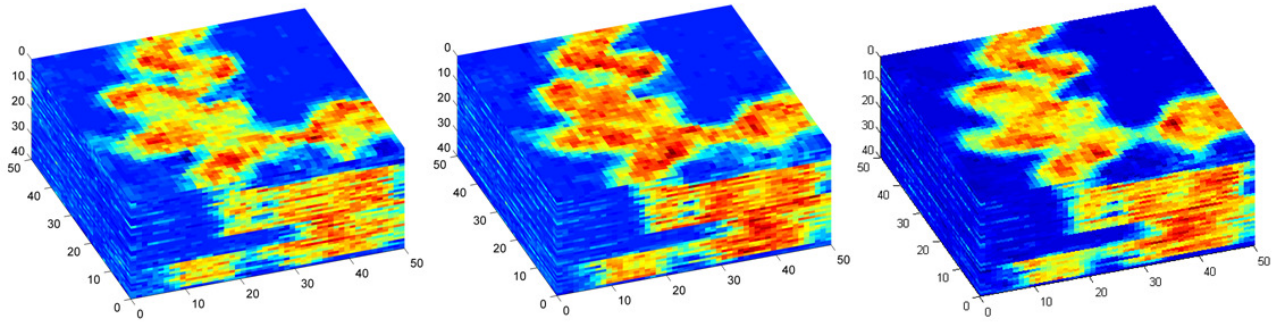

Figure 15: Samples from the posterior facies model under cluster $\mathcal{M}_1$.

computationally extensive run of ECLIPSE for $N_s = 7880$ times. Subsequently we use the Quasi Monte Carlo method to calculate the integral numerically.

Figure 18 shows the posterior probability of the model $\mathcal{M}_1$, $P(\mathcal{M}_1|\mathcal{D})$ using different orders of the polynomial chaos proxy. It is compared against the Quasi Monte Carlo method using the full reservoir simulator (ECLIPSE). It illustrates that the accuracy of the approximation increases as the order of expansion increases. It also demonstrates that the model $\mathcal{M}_1$ (training image Figure 7) is the most probable model among others. The plausibility of $\mathcal{M}_1$ increases as we allow smaller error in the observation. For example, the posterior probability of $\mathcal{M}_1$ is $P(\mathcal{M}_1|\mathcal{D}) = 0.942$ when the dimensionless error in the observation is $\sigma_D^2 = 0.1$, while $P(\mathcal{M}_1|\mathcal{D})$ decreases to 0.67 when the observational error increases to $\sigma_D^2 = 0.2$.

Figure 19 and Figure 20 illustrates the posterior probability of the model $\mathcal{M}_2$ and $\mathcal{M}_3$, $P(\mathcal{M}_2|\mathcal{D})$ and $P(\mathcal{M}_3|\mathcal{D})$ for different orders of the polynomial chaos expansion. It demonstrates that the model $\mathcal{M}_2$ is more plausible than the model $\mathcal{M}_3$. It also shows when the dimensionless observational error increases the plausibility of the model $\mathcal{M}_2$ and $\mathcal{M}_3$ increases. Figure 20 shows when the dimensionless observational error (σ_D) increases, the

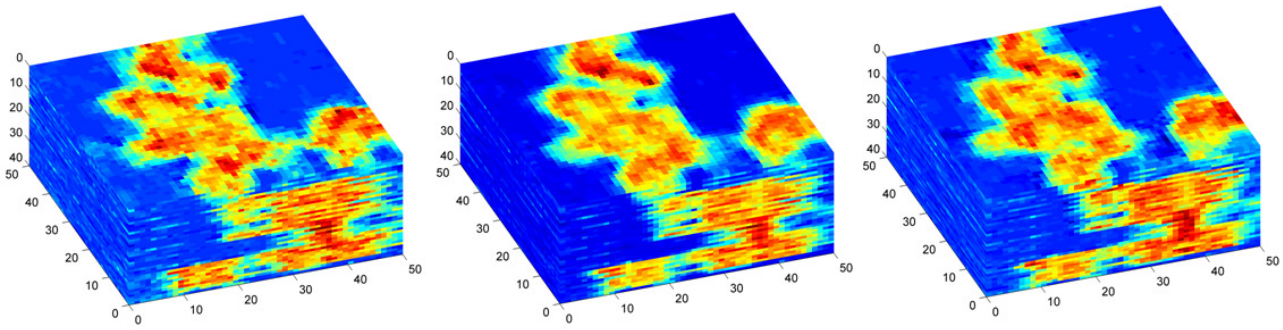

Figure 16: Samples from the posterior facies model under cluster $\mathcal{M}_2$.

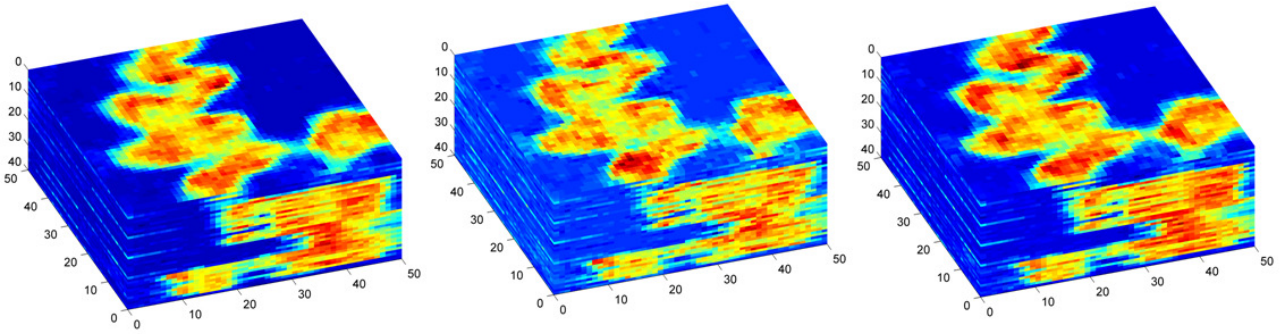

Figure 17: Samples from the posterior facies model under cluster $\mathcal{M}_3$.

misfit formulation becomes more nonlinear and higher order polynomial chaos have to be used for the accurate estimation of the posterior probability $P(\mathcal{M}_3|\mathcal{D})$.

Figure 21 shows the posterior probability of the model $\mathcal{M}_1$ when the number of retained eigenvalues increases. We used the polynomial chaos of order 6 for the proxy model. It demonstrates that the main difference between the probability of the models is picked up by the significant eigenvectors and the eigenvectors associated with smaller eigenvalues (in magnitude) do not contribute to the plausibility of the models remarkably. Accordingly, when the number of the retained eigenvectors increase from 10 to 40, $P(\mathcal{M}_1|\mathcal{D})$ does not change considerably. The dimensionless observational error for this experiment is assumed to be $\sigma_D^2 = 0.1$. Figure 22 shows the effect of increasing the number of eigenvectors on the posterior probability of the model $\mathcal{M}_2$.

Although in this specific case study the main difference between the plausibility of the models is picked up by the significant eigenvectors, the method proposed in this chapter is not limited to the case where there is a clear separation in the magnitude of the eigenvalues.

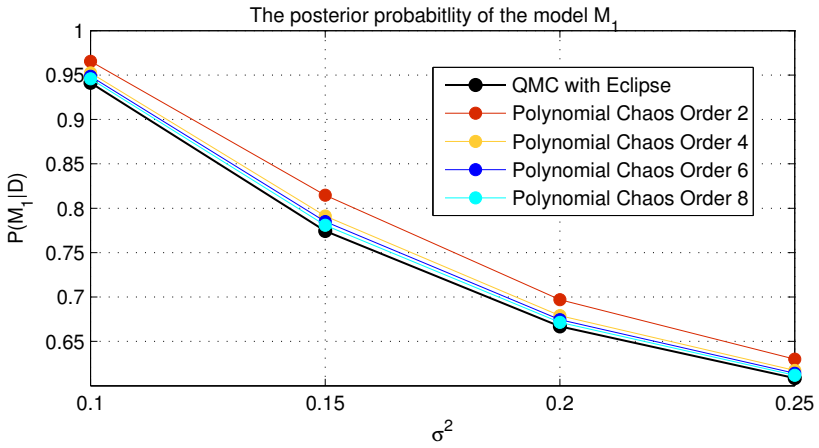

Figure 18: The posterior probability of the model $\mathcal{M}_1$, $P(\mathcal{M}_1|\mathcal{D})$, using different orders of the polynomial chaos proxy compared with the Quasi Monte Carlo method with ECLIPSE.

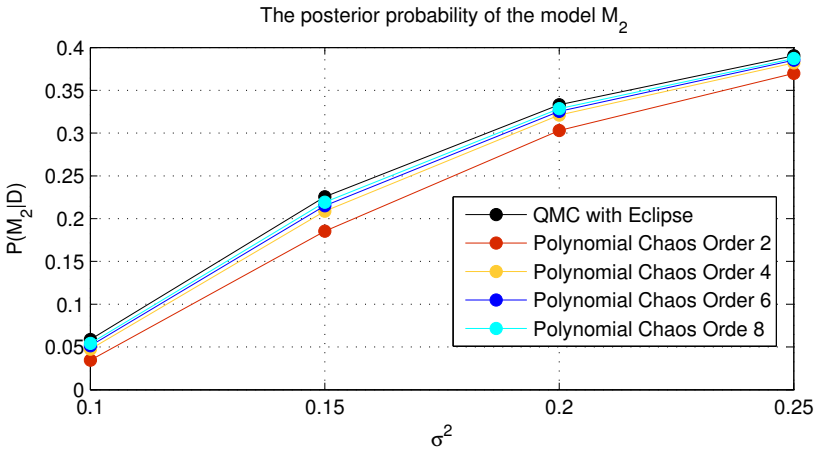

Figure 19: The posterior probability of the model $\mathcal{M}_2$, $P(\mathcal{M}_2|\mathcal{D})$, using different orders of the polynomial chaos proxy compared with the Quasi Monte Carlo method with ECLIPSE.

6. Concluding Remarks

In this work, we studied the "scenario uncertainty" where the prior model has a high level of discrete uncertainty. Such uncertainty usually comes from different interpretation of the sedimentary environments. For example, in real field applications, different scenarios of layering structure or fault interpretation may be plausible for a reservoir. The geological scenario usually has a considerable effect on flow response uncertainty. We use clustering approach to address the scenario uncertainty. Accordingly, we study the flow response uncertainty under each scenario and integrate the uncertainties to achieve the overall uncertainty. We applied the proposed cluster-based polynomial chaos proxy framework for an analytical Bayesian model selection problem and also to study the plausibility of three training images based on different geological interpretation of the second layer of synthetic Stanford VI reservoir, based on the given data. We demonstrated that the proposed workflow can be efficiently used to

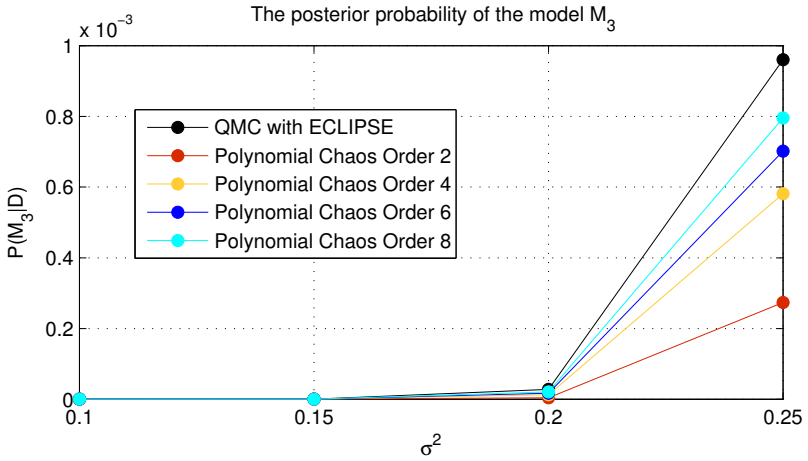

Figure 20: The posterior probability of the model $\mathcal{M}_3$, $P(\mathcal{M}_3|\mathcal{D})$, using different orders of the polynomial chaos proxy compared with the Quasi Monte Carlo method with ECLIPSE.

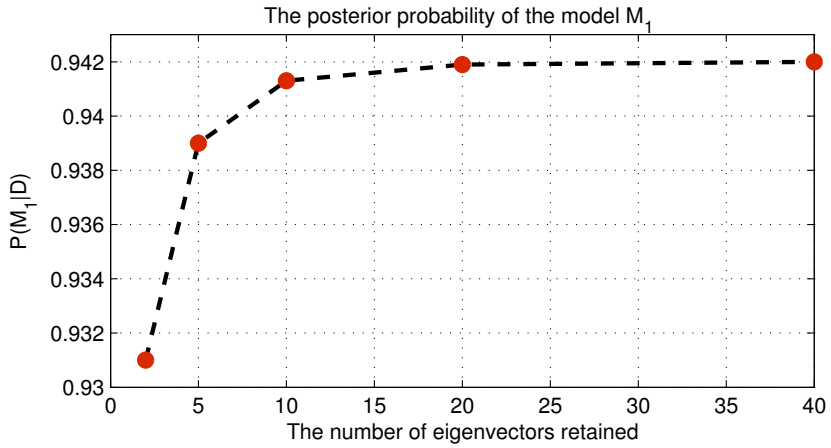

Figure 21: The posterior probability of the model $\mathcal{M}_1$, $P(\mathcal{M}_1|\mathcal{D})$, while different number of eigenvectors is retained.

calculate the posterior probability of each scenario and also sample from the posterior facies models within each scenario. The key findings of this work are as follows:

- To make the polynomial chaos representation sparse for the higher order terms, the polynomial chaos basis should be adapted to the input probability distribution, i.e. for the Gaussian input the Hermite polynomial basis gives an exponential convergence rate. However, in practice we are usually provided with only few realizations of the input distribution, hence it is not possible to construct the orthogonal polynomial basis adapted to the input distribution. Our solution to this problem is to reduce non-linearity by clustering, and make the polynomial chaos proxy sparse under each cluster.

- The first order polynomial chaos proxy is equivalent to the Laplace method to find

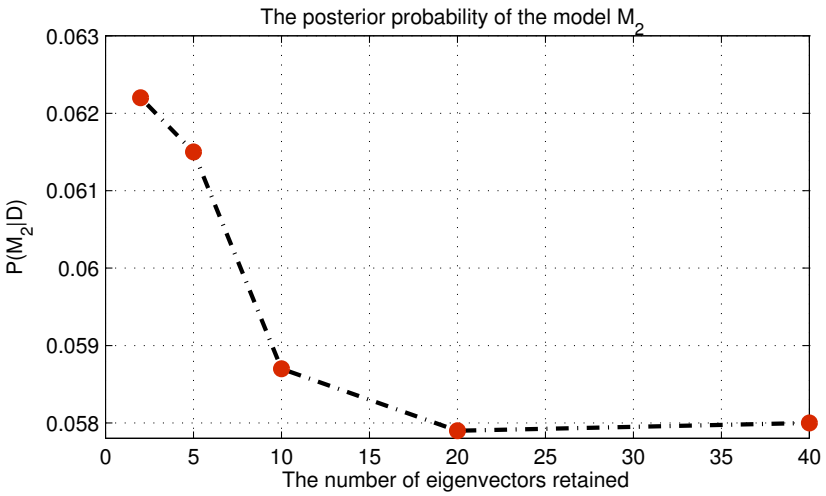

Figure 22: The posterior probability of the model $\mathcal{M}_2$, $P(\mathcal{M}_2|\mathcal{D})$, while different number of eigenvectors is retained.

the Bayes factor. The higher order terms are demonstrated to be more accurate than the Laplace method. The reason is that the Laplace method does not capture the nonlinear terms in the probability distribution function and approximate it with a Gaussian distribution around the maximum likelihood point.

- When the prior clustering quality is poor, i.e. the variance of the realizations under each cluster is high, the number of eigenvalues required to be retained to accurately represent the input realizations increases. Hence, the number of relevant terms in the reduced terms polynomial chaos expansion increases accordingly. For a given number of trial runs, this reduces the maximum number of the terms could be accommodated in the polynomial chaos approximation and accordingly the accuracy decreases. Similarly, when the realizations under each cluster do not share the same statistics, the convergence rate of the polynomial chaos proxy becomes slow and higher-order terms are required to accurately represent the input realizations.

- In general, the accuracy of the cluster-based polynomial chaos proxy depends on:

  1. The quality of the trial simulation runs and the input realizations
  2. The order of the polynomial chaos proxy under each cluster
  3. The number of eigenvalues retained
  4. The quality of the clustering

References

[1] Bazargan, Hamid, et al. *Surrogate accelerated sampling of reservoir models with complex structures using sparse polynomial chaos expansion*, Advances in Water Resources **86** (2015): 385-399.

[2] Maurice George Kendall, *Multivariate analysis*, Macmillan New York, **1980**.

[3] Bahar Erar, *Mixture model cluster analysis under different covariance structures using information complexity*, **2011**.

[4] Mark Leonard Davison, *Multidimensional scaling*, Wiley New York, **1983**.

[5] Joseph B Kruskal, *Nonmetric multidimensional scaling: a numerical method*, Psychometrika **29 (1964)**, no. 2, 115129.

[6] Leonard Kaufman and Peter J Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, vol. 344, Wiley-Interscience, **2009**.

[7] Michael R Anderberg, *Cluster analysis for applications*, Tech. report, DTIC Document, **1973**.

[8] Stephen C Johnson, *Hierarchical clustering schemes*, Psychometrika **32 (1967)**, no. 3, 241 254.

[9] Kardi Teknomo, *K-means clustering tutorial*, Medicine **100 (2006)**, no. 4, 3.

[10] Markus Svensen and Christopher M Bishop, *Robust Bayesian mixture modelling, Neurocomputing* **64 (2005)**, 235252.

[11] Harold Jeffreys, *Theory of probability*, Oxford University Press, **1998**.

[12] DR Cox, *Tests of separate families of hypotheses*, **1961**

[13] Hirotugu Akaike, *A new look at the statistical model identification*, Automatic Control, IEEE Transactions on **19 (1974)**, no. 6, 716723

[14] Gideon Schwarz, *Estimating the dimension of a model*, The annals of statistics **6 (1978)**, no. 2, 461464.

[15] Merlise Clyde and Edward I George, *Model uncertainty*, Statistical science **(2004)**, 8194.

[16] Christophe Andrieu and Arnaud Doucet, *Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC*, Signal Processing, IEEE Transactions on **47 (1999)**, no. 10, 26672676.

[17] Christophe Andrieu, Nando De Freitas, and Arnaud Doucet, *Sequential MCMC for Bayesian model selection*, Higher-Order Statistics, **1999**. Proceedings of the IEEE Signal Processing Workshop on, IEEE, 1999, pp. 130134.

[18] Hugh Chipman, Edward I George, Robert E McCulloch, M Clyde, Dean P Foster, and Robert A Stine, *The practical implementation of Bayesian model selection*, Lecture Notes Monograph Series **(2001)**, 65134.

[19] Bradley P Carlin and Siddhartha Chib, *Bayesian model choice via Markov chain Monte Carlo methods*, Journal of the Royal Statistical Society. Series B (Methodological) **(1995)**, 473484.

[20] Peter J Green, *Reversible jump Markov chain Monte Carlo computation and bayesian model determination*, Biometrika **82 (1995)**, no. 4, 711732.

[21] Robert E Kass and Adrian E Raftery, *Bayes factors*, Journal of the american statistical association **90 (1995)**, no. 430, 773795.

[22] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski, *Using Bayesian model averaging to calibrate forecast ensembles*, Monthly Weather Review **133 (2005)**, no. 5, 11551174.

[23] Yulia Gel, Adrian E Raftery, and Tilmann Gneiting, *Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method*, Journal of the American Statistical Association **99 (2004)**, no. 467, 575583.

[24] Kerry Gallagher, Karl Charvin, Soren Nielsen, Malcolm Sambridge, and John Stephenson, *Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for earth science problems*, Marine and Petroleum Geology **26 (2009)**, no. 4, 525535.

[25] Geoffrey J. McLachlan, RW Bean, and David Peel, *A mixture model-based approach to the clustering of microarray expression data*, Bioinformatics **18 (2002)**, no. 3, 413422.

[26] Brian S Everitt and David J Hand, *Finite mixture distributions*, Monographs on Applied Probability and Statistics, London: Chapman and Hall, **1 (1981)**.

[27] Edward W Forgy, *Cluster analysis of multivariate data: efficiency versus interpretability of classifications*, Biometrics **21 (1965)**, 768769

[28] James MacQueen et al., *Some methods for classification and analysis of multivariate observations*, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, California, USA, 1967, p. 14.

[29] Steven N Goodman, *Toward evidence-based medical statistics. 2: The Bayes factor*, Annals of internal medicine **130 (1999)**, no. 12, 10051013.

[30] David Madigan and Adrian E Raftery, *Model selection and accounting for model uncertainty in graphical models using Occams window*, Journal of the American Statistical Association **89 (1994)**, no. 428, 15351546.

[31] Bradley P Carlin and Siddhartha Chib, *Bayesian model choice via Markov chain Monte Carlo methods*, Journal of the Royal Statistical Society. Series B (Methodological) **(1995)**, 473484.

[32] Simon J Godsill, *On the relationship between MCMC model uncertainty methods*, Citeseer, **1998**.

[33] Sylvia Richardson and Peter J Green, *On Bayesian analysis of mixtures with an unknown number of components (with discussion)*, Journal of the Royal Statistical Society: series B (statistical methodology) **59 (1997)**, no. 4, 731792.

[34] C Andrieu, A Doucet, and CP Robert, *Computational advances for and from Bayesian analysis*, Statistical science **19 (2004)**, no. 1, 118127.

[35] Luke Tierney and Joseph B Kadane, *Accurate approximations for posterior moments and marginal densities*, Journal of the American Statistical Association **81 (1986)**, no. 393, 8286.

[36] Hugh Chipman, Edward I George, Robert E McCulloch, M Clyde, Dean P Foster, and Robert A Stine, *The practical implementation of Bayesian model selection*, Lecture Notes Monograph Series **(2001)**, 65134.

[37] Robert McCulloch and Peter E Rossi, *An exact likelihood analysis of the multinomial probit model*, Journal of Econometrics **64 (1994)**, no. 1, 207240.

[38] Larry Wasserman, *Asymptotic inference for mixture models by using data-dependent priors*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **62 (2000)**, no. 1, 159180.

[39] Arthur E Hoerl and Robert W Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics **12 (1970)**, no. 1, 5567.

[40] Robert Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society. Series B (Methodological) **(1996)**, 267288.

[41] David Draper, *Assessment and propagation of model uncertainty*, Journal of the Royal Statistical Society. Series B (Methodological) **(1995)**, 4597.

[42] Edward E Leamer, *Specification searches: ad hoc inference with nonexperimental data*, Wiley New York, **1978**.

[43] Edward I George, *The variable selection problem*, Journal of the American Statistical Association **95 (2000)**, no. 452, 13041308.

[44] Chris Chatfield, *Model uncertainty*, Encyclopedia of Environmetrics **(1995)**.

[45] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry S Davis, *Background and foreground modeling using nonparametric kernel density estimation for visual surveillance*, Proceedings of the IEEE **90 (2002)**, no. 7, 11511163

[46] C Scheidt and J Caers, *A new method for uncertainty quantification using distances and kernel methods. Application to a deepwater turbidite reservoir*, SPEJ, SPE-118740-PA **(2009)**.

[47] Celine Scheidt and Jef Caers, *Uncertainty quantification in reservoir performance using distances and kernel methodsapplication to a West Africa deepwater turbidite reservoir*, SPE Journal **14 (2009)**, no. 4, 680692.

[48] Temstocles Rojas, Vasily Demyanov, Mike Christie, and Dan Arnold, *Reducing uncertainty in modelling fluvial reservoirs by using intelligent geological priors*, Geostatistical Congress, Olso, **2012**.

[49] Simon Tong and Daphne Koller, *Support vector machine active learning with applications to text classification*, The Journal of Machine Learning Research **2 (2002)**, 4566.

[50] Hyucksoo Park, Celine Scheidt, Darryl Fenwick, Alexandre Boucher, and Jef Caers, *History matching and uncertainty quantification of facies models with multiple geological interpretations*, Computational Geosciences **(2013)**, 113.

[51] S Castro, J Caers, and T Mukerji, *The stanford vi reservoir, Stanford Center for Reservoir Forecasting (SCRF)*, Annual Report **(2005)**.

[52] Sebastien Strebelle, *Conditional simulation of complex geological structures using multiple-point statistics*, Mathematical Geology **34 (2002)**, no. 1, 121

[53] Harald Niederreiter, *Low-discrepancy and low-dispersion sequences*, J. Number Theory **30 (1988)**, no. 1, 5170.

[54] Ian H Sloan and Henryk Wozniakowski, *When are quasi-monte carlo algorithms efficient for high dimensional integrals?*, Journal of Complexity **14 (1998)**, no. 1, 133.

[55] Russel E Caflisch, *Monte carlo and quasi-monte carlo methods*, Acta numerica **1998, 149**

[56] Erich Novak and Klaus Ritter, *High dimensional integration of smooth functions over cubes*, Numerische Mathematik **75 (1996)**, no. 1, 7997.

[57] Thomas Gerstner and Michael Griebel, *Numerical integration using sparse grids*, Numerical algorithms **18 (1998)**, no. 3-4, 209232.

[58] Ilya M Sobol, *On quasi-monte carlo integrations*, Mathematics and Computers in Simulation **47 (1998)**, no. 2, 103112.

[59] Adrian E Raftery, David Madigan, and Jennifer A Hoeting, *Bayesian model averaging for linear regression models*, Journal of the American Statistical Association **92 (1997)**, no. 437, 179191

[60] Wiener, Norbert. *The homogeneous chaos.* American Journal of Mathematics **60.4 (1938)**: 897-936.

[61] R. Cameron and W. Martin, *The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals*, Ann. Math., **48 (1947)**, p. 385.